DOCUMENT RESUME

ED 069 788 TM 002 274

AUTHOR	Elashoff, Janet Dixon; Dunbar, Charles R.
TITLE	Measures of Association.
INSTITUTION	Stanford Univ., Calif. Stanford Center for Research and Development in Teaching.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C.
BUREAU NO	BR-5-0252
PUB DATE	Aug 72
CONTRACT	OEC-6-10-078-9C
NOTE	51p.; R&D Memo No. 93
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Comparative Analysis; *Correlation; *Educational Research; *Measurement Instruments; *Measurement Techniques; Statistical Data

ABSTRACT

A common problem in educational research is measuring the degree of relationship or association between two variables. Many investigators habitually use Pearson's product-moment correlation coefficient or a transformation of x2. In the past two decades, however, a variety of association measures have been introduced in the statistics literature. This report contains a review of available association measures, supplemented by discussion of the several factors involved in selecting a measure of association such as the types of variables (continuous, ranked, ordered) and the type of association expected (linear, monotone, general). Examples illustrate the necessary calculations and provide comparisons among the measures. (Author)

NCERD
BR-5-0252

TM

SC
RD
T

Stanford Center for Research and Development in Teaching

SCHOOL OF EDUCATION STANFORD UNIVERSITY

ED 069788

STANFORD CENTER
FOR RESEARCH AND DEVELOPMENT
IN TEACHING

Research and Development Memorandum No. 93

MEASURES OF ASSOCIATION

Janet Dixon Elashoff and Charles R. Dunbar

School of Education
Stanford University
Stanford, California

August 1972

Published by the Stanford Center for Research and Development in Teaching, supported in part as a research and development center by funds from the United States Office of Education, Department of Health, Education, and Welfare. The opinions expressed in this publication do not necessarily reflect the position or policy of the Office of Education and no official endorsement by the Office of Education should be inferred. (Contract No. OEC-6-10-078, Component 9C.)

Introductory Statement

The Center's mission is to improve teaching in American schools. Too many teachers still employ a didactic style aimed at filling passive students with facts. The teacher's environment often prevents him from changing his style, and may indeed drive him out of the profession. And the children of the poor typically suffer from the worst teaching.

The Center uses the resources of the behavioral sciences in pursuing its objectives. Drawing primarily upon psychology and sociology, but also upon other behavioral science disciplines, the Center has formulated programs of research, development, demonstration, and dissemination in three areas. Program 1, Teaching Effectiveness, is now developing a Model Teacher Training System that can be used to train both beginning and experienced teachers in effective teaching skills. Program 2, The Environment for Teaching, is developing models of school organization and ways of evaluating teachers that will encourage teachers to become more professional and more committed. Program 3, Teaching Students from Low-Income Areas, is developing materials and procedures for motivating both students and teachers in low-income schools.

Many research studies result in the computation of measures of association between pairs of variables. This paper provides a review of the variety of measures available and explains the circumstances under which each is appropriate. This paper, then, should help educational researchers make use of appropriate statistical methodology in studying relationships between variables.

Table of Contents

List of Tables

List of Figures

Abstract

A common problem in educational research is measuring the degree of relationship or association between two variables. Many investigators habitually use Pearson's product-moment correlation coefficient or a transformation of χ^2. In the past two decades, however, a variety of association measures have been introduced in the statistics literature. This report contains a review of available association measures, supplemented by discussion of the several factors involved in selecting a measure of association, such as the types of variables (continuous, ranked, ordered) and the type of association expected (linear, monotone, general). Examples illustrate the necessary calculations and provide comparisons among the measures.

MEASURES OF ASSOCIATION

Janet Dixon Elashoff and Charles R. Dunbar

Introduction

A common problem in educational research is that of measuring the degree of relationship or association between two variables. For example, an investigator might wish to estimate the degree to which achievement scores can be predicted from IQ scores or the degree of agreement between two raters using the same five-point scale. A wide variety of association measures is available--Pearson's r, Kendall's τ, Goodman and Kruskal's γ, and others. This study reviews some association measures discussed in the statistical literature and offers guidelines for making an appropriate choice of measure for a particular problem.

Each measure of association was developed to be applied in a particular class of problems. Thus, in our review we pay special attention to the several major factors determining the type of problem for which a measure was designed:

1. What type of measurement scale do the two variables represent?
2. Does one variable logically precede the other; that is, will one variable be used to predict the other and should the measure of association reflect this?
3. What type of relationship between the two variables is the measure sensitive to?

4. What sampling conditions and assumptions about the joint distribution of the two variables are necessary for standard tests of significance to be valid?

For this study, association measures have been grouped according to the type of measurement scale for which they are designed. Variables such as age or height are designated as continuous variables even though they are usually rounded off to the nearest year or inch. Discrete variables are classified into four basic types: (a) rank-ordered values, observations ranked from 1 to n; (b) ordered multicategorical values, observations assigned scores such as 1, 2, 3, 4, 5; (c) unordered nominal values; (d) dichotomous values. Although dichotomous variables can logically be included in type (b) or (c), some measures of association have been developed specifically for them.

Naturally, a variable that is intrinsically continuous could be turned into a ranked or ordered categorical variable, or for some purposes a variable of type (a), (b), or (d) could be treated as continuous. Therefore, classification of variables by this scheme may be somewhat arbitrary and should serve mainly as a preliminary guide to choosing a measure of association. The final choice should rest most heavily on consideration of the type of relationship between variables that is of interest.

Measures of association are intended to describe the degree of relationship between two variables and are usually defined to be +1.0 (or -1.0) for a perfect predictive relationship and 0.0 for no relationship. Each measure of association is designed for a different type of relationship. For example, since Pearson's r is designed to measure the degree of linear relationship, $r = 1$ only for perfect linear relationships, and $r = 0$ cannot be used to infer the "independence" of the two variables in the population; it merely indicates no linear relationship in the sample. Many different kinds of relationships are possible between two variables: (a) a linear relationship--the relationship between a pretest and a posttest IQ score might be expected to be linear; (b) a monotone relationship--e.g., average weight increases with height, however, the average difference in weight for two inches' difference in height may be different at heights near 30 inches than at heights near 72 inches; (c) general association--e.g., small-group discussions occur much more frequently in connection with social studies lessons, whereas individual work is more often associated with math lessons.

In selecting a measure of association it is most important to have in mind the type of relationship that might be expected to occur or that would be of most interest. A first step in this selection procedure is to arrange the data in a scatterplot or two-way table so that a visual assessment can be made of the relationship between the two variables and of any peculiarities in the data that might invalidate the choice of a particular measure of association.

Some measures of association, such as Pearson's r, are said to be symmetric in the two variables. That is, regardless of whether the prediction is x from y or y from x, the measure of association, r, is the same. Other measures, like the correlation ratio η^2 or the measure λ, are said to be asymmetric; that is, under these models, prediction of y from x might be more accurate than prediction of x from y, and thus

$\eta^2_{y \cdot x} \neq \eta^2_{x \cdot y}$ in general. Unless otherwise noted, measures of association are usually symmetric in the two variables.

Measures of association can be defined for a population or a sample. For brevity we give only the formula for the sample measure; the reader interested in the corresponding population measures should refer to the literature cited for each measure. If a sample has been selected at random from some population, then the sample measure may be used to make inferences about the value of the population measure. This assumption of random sampling is implicit in all tests of the statistical significance of a measure of association. Other assumptions necessary for testing the statistical significance of an observed degree of association will be discussed in conjunction with each particular measure.

Continuous or Rank-Ordered Variables

If both variables are continuous, the standard measure of association is Pearson's r. Two nonparametric measures, Spearman's r and Kendall's tau are also applicable in this case. Two hypothetical examples will be used to illustrate these measures.

In a research study, several teachers were observed in their classrooms over a six-week period. Among other things, the observers coded negative feedback (e.g., criticism, rebuke) given by these teachers to 18 randomly chosen students, and, at the end of the time period, a self-esteem assessment scale was administered to the students. The data are shown in Table 1 and Figure 1.

For the second example, shown in Table 2 and Figure 2, an instructor in an educational psychology course wished to use the class results on

TABLE 1

Student Self-Esteem and Rate of Negative Feedback from the Teacher (Average Negative Responses per Pupil Classroom Hour)

Student	Self-esteem	Negative-feedback rate	Student	Self-esteem	Negative-feedback rate
1	44	4.13	10	64	2.75
2	70	3.38	11	100	1.25
3	110	.38	12	34	5.88
4	42	5.63	13	74	2.50
5	68	1.88	14	32	7.00
6	88	2.36	15	120	.64
7	72	2.12	16	62	3.82
8	90	1.00	17	56	3.00
9	102	.85	18	86	1.50

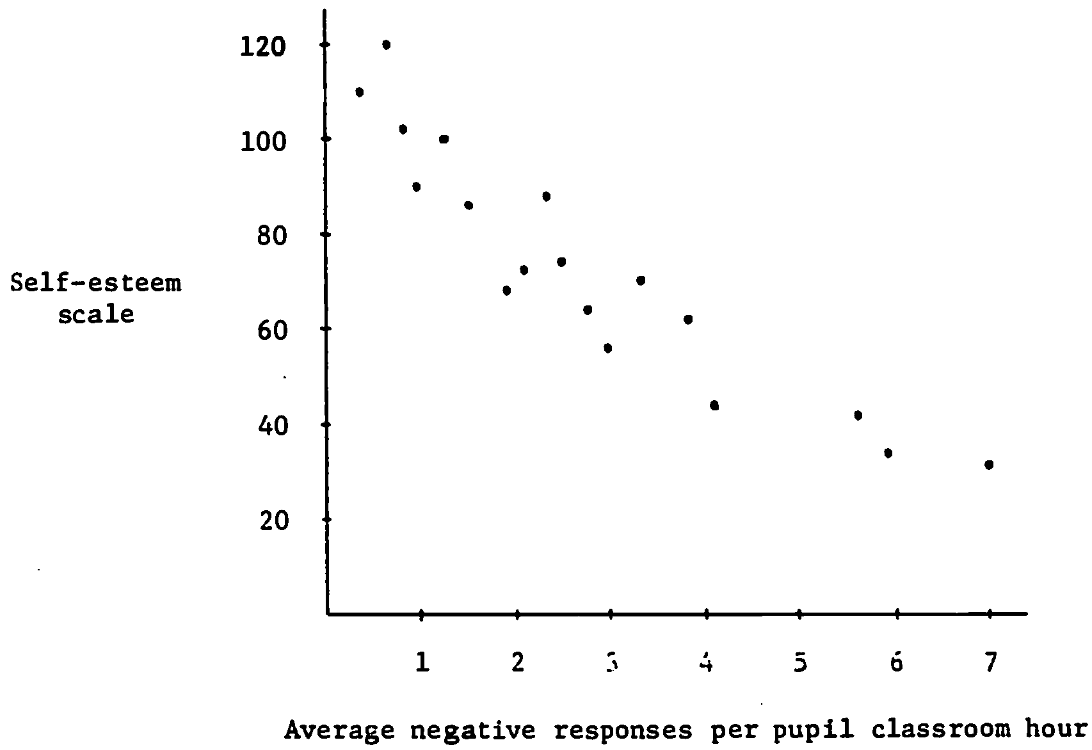


Fig. 1. Scatterplot of self esteem vs. negative-feedback rate.

TABLE 2

Scholastic Aptitude Subtest Results for Twenty-Four Students
(Possible Score of 800)

Student	Verbal	Mathematical	Student	Verbal	Mathematical
1	770	690	13	800	770
2	540	480	14	740	790
3	610	540	15	570	660
4	630	510	16	680	720
5	590	640	17	580	610
6	700	540	18	660	490
7	650	530	19	450	400
8	510	380	20	610	460
9	610	680	21	560	550
10	520	420	22	620	680
11	690	590	23	610	680
12	670	760	24	660	650

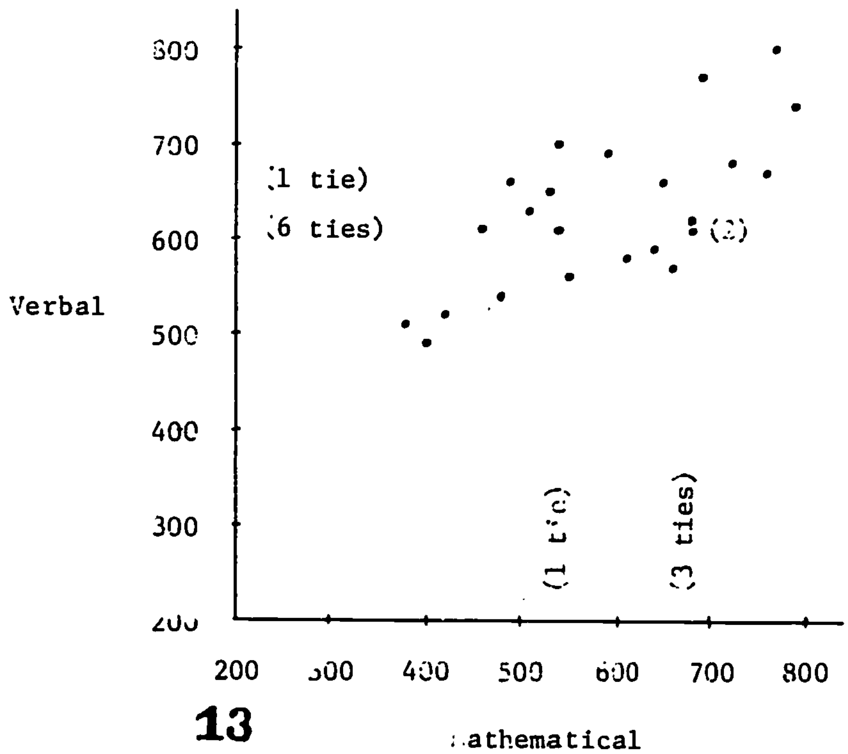


Fig. 2. Scatterplot of verbal vs. mathematical SAT scores.

the subtests of the Scholastic Aptitude Test to demonstrate that the two test sections usually have a correlation near .64.

The observations on the two variables will be denoted x_i and y_i for individuals i = 1, ..., n.

Pearson Product-Moment Correlation

The most commonly used measure of association for two continuous variables is the Pearson product-moment correlation. The population measure is usually denoted by ρ and the sample measure by r. Pearson's r is designed for situations in which the relationship between variables x and y is linear. The measures ρ and r will be +1.0 for a perfect linear relationship with positive slope (see Figure 3a) and −1.0 for a perfect linear relationship with negative slope (Figure 3b). They will be zero if there is no linear relationship (see Figure 3 c and d). The population measure ρ is defined as

$$(1) \qquad \rho = \frac{\text{Covariance }(x,y)}{\sigma_x \sigma_y} = \frac{E(X - \mu_x)(Y - \mu_y)}{\left[E(X - \mu_x)^2 \; E(Y - \mu_y)^2\right]^{1/2}}$$

The sample statistic r is

$$(2) \qquad r = \frac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{\left[\sum_i (x_i - \overline{x})^2 \cdot \sum (y_i - \overline{y})^2\right]^{1/2}}$$

For the self-esteem example from Table 1, r = −.924. This means there is a close linear relationship between negative-feedback rate and self-esteem, but in a negative direction. As negative responses increase,

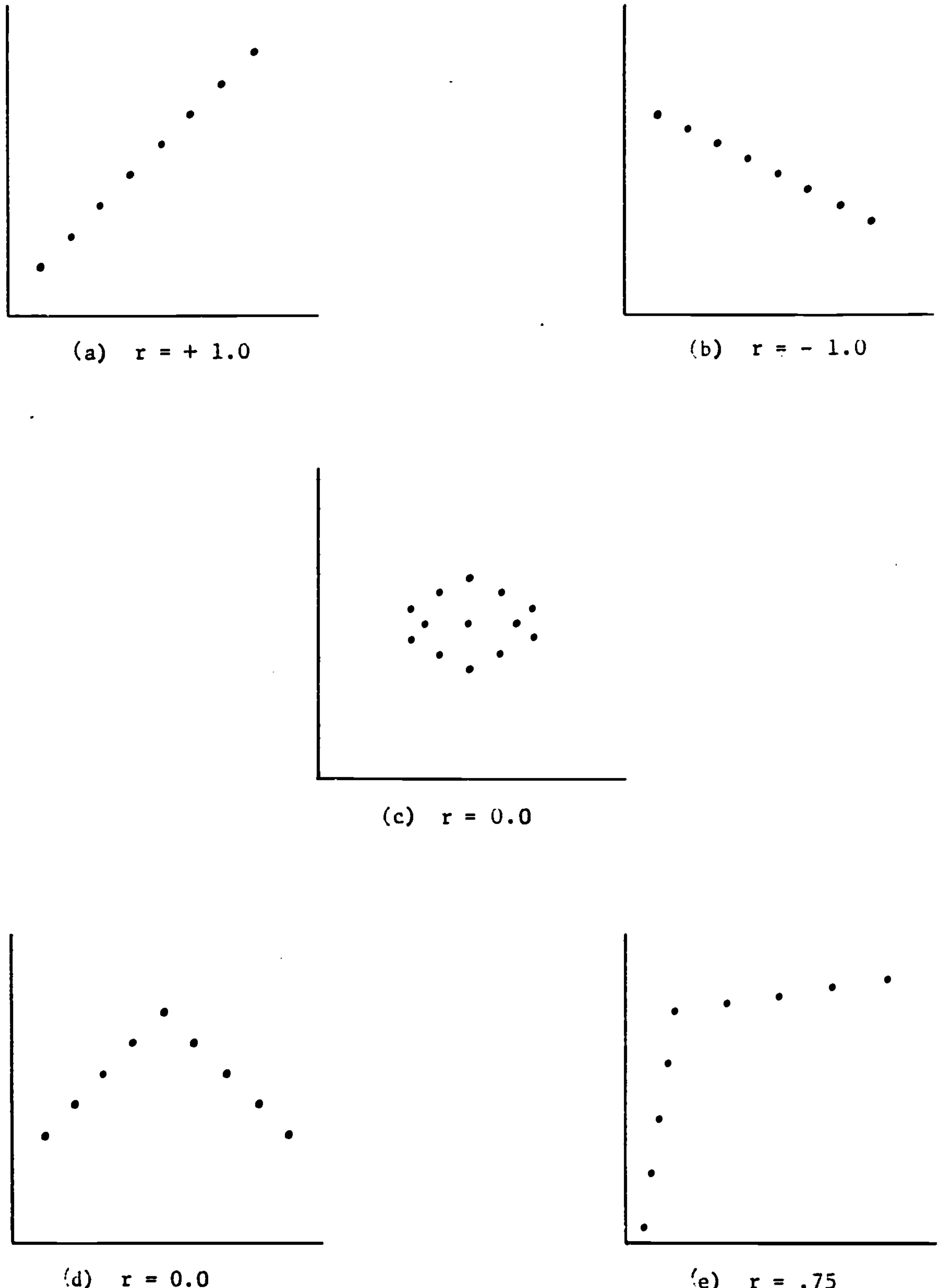


Fig. 3. Scatterplots showing different values of r.

self-esteem decreases. For the data in Table 2, $r = .692$, representing a fairly high, positive degree of linear relationship. It is also well within the instructor's anticipated limits.

To make inferences about the population measure ρ, using r, it must be assumed that the sample observations were randomly selected from the population and that x and y have a bivariate normal distribution. Bivariate normality implies that both x and y have normal distributions and that the relationship between x and y is linear. For tests and confidence intervals about the value of ρ, see, e.g., Dixon and Massey (1969) or Hays (1963). For discussions of the many factors affecting the size of r, see Carroll (1961) or Walker and Lev (1969).

In cases where the two variables can be assumed to have a bivariate normal distribution and for some reason one or both variables have been dichotomized but the investigator still wishes to make inferences about the population value of ρ, the measures tetrachoric r, point biserial r, ϕ, and biserial r have been developed. These are discussed later in in this study. When the variables have been categorized into more than two categories, estimation methods for ρ using the polychoric series method have been developed by Lancaster and Hamdan (1964).

Pearson's r is designed for situations in which the relationship between variables is linear, and for inferences about ρ to be valid the joint distribution of x and y must be bivariate normal. If a monotone, but not necessarily linear, relationship is of interest, or bivariate normality is unlikely, the nonparametric measures of association, Spearman's rank correlation coefficient or Kendall's tau, should be considered.

Spearman's Rank-Correlation Coefficient

Spearman's r is designed to measure the degree of monotone relationship between two variables x and y. Instead of using the exact score values, the observations are ranked from lowest to highest on each variable separately and then Pearson's r is calculated on the ranks. When two variables have a monotone relationship, their ranks will have a linear relationship.

For convenience, the mathematically equivalent computing formula

$$r_s = 1 - \frac{6\sum D_i^2}{n(n^2 - 1)} \qquad (3)$$

may be used. The number D_i is the difference between the x and y ranks for the i^{th} individual.

Spearman's r will be +1.0 for perfect positive monotone relationships such as those shown in Figure 3 a and e, -1.0 for perfect negative monotone relationships, and 0.0 if there is no relationship or the relationship is not monotone (Figure 3 c or d).

For the self-esteem data shown in Table 1, the rank scores are shown in Table 3 and the rank scatterplot in Figure 4. The obtained $r_s = -.948$ is very close to the Pearson's r = -.924.

Sometimes observations will be tied as in the mathematical and verbal scores from Table 2. If the number of ties is small, the midranks of the tied observations can be used (see below) and formula (3) can still be applied. If the number of tied observations is large, then the reader should refer to the formula given in Kendall (1970).

The rankings for the data of Table 2 are found in Table 4 and Figure 5. Midranks are assigned by averaging the rank positions which tied

TABLE 3

Ranked Scores on x (Negative Feedback) and y (Self-Esteem) for Eighteen Students Arranged in Order by Rank on x

Student	x ranks	y ranks	D_i
14	1	18	-17
12	2	17	-15
4	3	16	-13
1	4	15	-11
17	5	12	-7
16	6	14	-8
10	7	11	-4
5	8	7	1
2	9	13	-4
7	10	8	2
13	11	16	1
18	12	6	6
6	13	9	4
8	14	4	10
11	15	5	10
9	16	3	13
3	17	1	16
15	18	2	16

$$r_s = 1 - \frac{6\sum D_i^2}{n(n^2 - 1)} = 1 - \frac{11328}{5814} = 1 - 1.948 = - .948$$

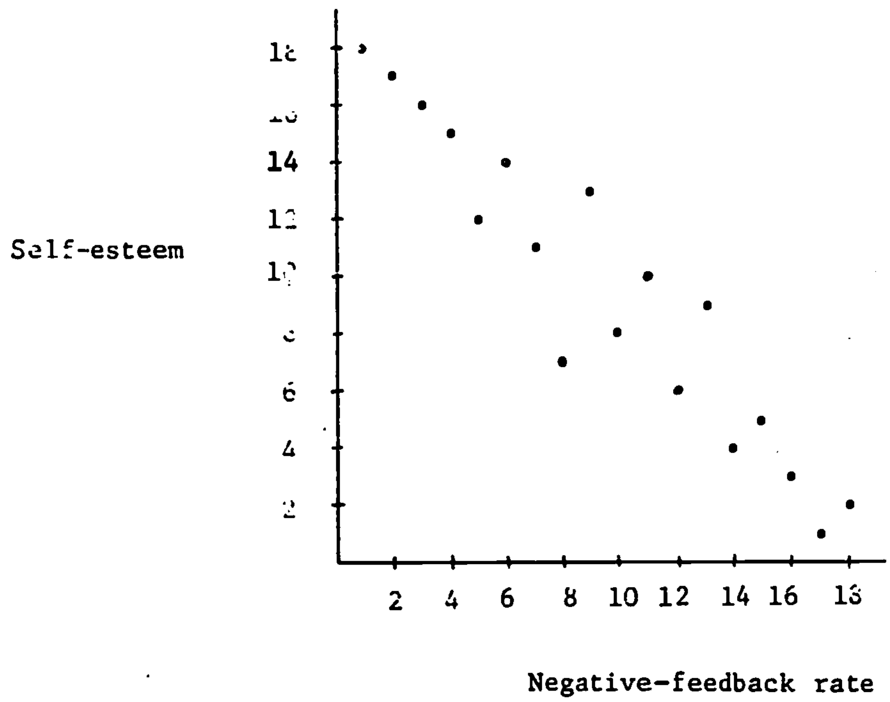


Fig. 4. Rank scatterplot for self-esteem data.

TABLE 4

Twenty-four Students Ordered by Rank on the Mathematics Subtest x

Student	x ranks (mathematical subtest)	y ranks (verbal)
14	1	3
13	2	1
12	3	7
16	4	6
1	5	2
9	7.0	14.5
22	7.0	12
23	7.0	14.5
15	9	19
24	10	8.5
5	11	17
17	12	18
11	13	5
21	14	20
3	15.5	14.5
6	15.5	4
7	17	10
4	18	11
18	19	8.5
2	20	21
20	21	14.5
10	22	22
19	23	24
8	24	23

$$r_s = 1 - \frac{6(832.5)}{24(24^2 - 1)} = 1 - \frac{4995}{13800} = .638$$

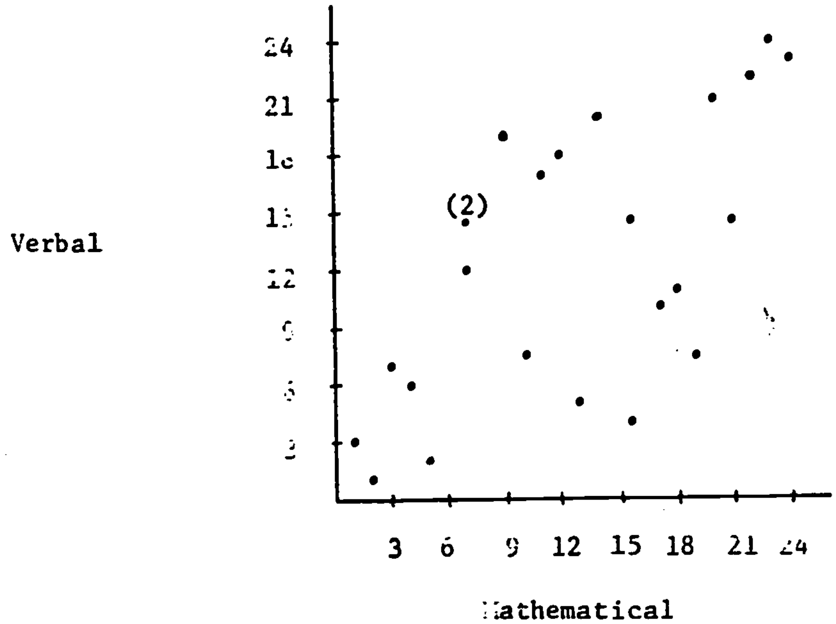


Fig. 5. Rank scatterplot of SAT data.

observations would have if they were distinguishable. In Table 3, students 9, 22, and 23 are tied on mathematics (x) scores. If their scores were distinguishable, these three would hold ranks 6, 7, and 8. Since they are tied, they are all assigned the midrank $(6 + 7 + 8)/3 = 7.0$. Similarly, students 3 and 6 would hold ranks 15 and 16, but because they are tied, both are given rank $(15 + 16)/2 = 15.5$. For the ranks of Table 4, $r_s = .638$ as compared to Pearson's $r = .692$.

If the observations represent a random sample from some population, the null hypothesis that x and y do not have a monotone relationship (that is, that their ranks do not have a linear relationship) can be tested. Use of r_s to make inferences about association in the population requires that observations can be ranked without ties, so x and y must be continuous variables. For small n ($n \leq 30$), tables of the critical values of r_s may be found in Siegel (1956). For large n, tables and tests for Pearson's r provide approximate tests of the significance of r_s.

For measures of association designed for continuous variables, the occurence of ties can affect the validity of the significance tests. If the proportion of ties is small, the method outlined above should be reasonable. Other approaches to the handling of ties are possible, for example, see the section on "ambiguous data" in Bradley (1968). If the proportion of ties is not small, one of the association measures designed for categorical variables should be considered instead.

Kendall's Tau

Like Spearman's r_s, Kendall's tau, τ, is a nonparametric rank measure of association which measures the degree of monotone relationship between two variables; τ, however, is derived from different principles.

Kendall's τ is +1.0 for a perfect positive monotone relationship between x and y, −1.0 for a negative relationship, and 0.0 when no monotone relationship exists. The measure τ is based on the idea of examining all possible pairs of observations and recording for each whether the relative ranks assigned to x agree with those assigned to y. For example, for students numbered 14 and 12 from Table 3, the x ranks are 1 and 2 and the y ranks are 18 and 17 respectively. Thus, these two individuals are ranked in the opposite order by variables x and y and provide an instance of disagreement between the two rankings. Students numbered 17 and 16 provide an instance of agreement between the two rankings. After all pairs of observations are examined, the difference between the total number of agreements (P) and the total number of disagreements (Q) is compared to the maximum possible number of agreements $(n(n-1))/2$. Thus,

$$(4) \qquad \tau = \frac{P - Q}{\frac{n(n-1)}{2}}$$

Consider the data from Table 3. Since there are 18 cases, the total number of pairs is $\binom{n}{2} = \frac{n!}{(n-2)!2!} = \frac{n(n-1)}{2} = \frac{18 \cdot 17}{2} = 153$. When there are no ties it is only necessary to count P or Q, but not both, because $P + Q = \frac{n(n-1)}{2}$. That is, the number of pairs whose ranks disagree and the number of pairs whose ranks agree must add up to the total number of pairs. In the example, the student pairs numbered (17,16), (17,2), (10,2), (5,2), (5,13), (5,6), (7,6), (18,6), (8,11), (3,15) agree in x and y rank order. Thus P = 11 and it follows that Q = 153 − 11 = 142. The result is $\tau = \frac{11 - 142}{153} = \frac{-131}{153} = -.856$. This indicates a high degree of negative monotonicity; note that τ is somewhat smaller than Spearman's rank correlation or the Pearson product-moment correlation.

Assuming that the individuals represent a random sample from some population in which x and y are continuous variables, τ may be used to test the null hypothesis of no monotone relationship between x and y. The test is based on the asymptotic normality of $S = P - Q$. For large n, the variance of S is

(5) $$\sigma_S^2 = \frac{1}{18} n(n-1)(2n+5)$$

Thus if x and y have no monotone relationship,

$$z = \frac{S}{\sigma_S}$$

has approximately a standard normal distribution, and the null hypothesis would be rejected if $P - Q > \sigma_S z_{1-\alpha/2}$ or $P - Q < \sigma_S z_{\alpha/2}$. See Kendall (1970) for additional details and an explanation of the continuity correction that should be used when n is small.

Kendall's tau requires that the variables be continuous so that no ties in the ranks can occur. Because of measurement problems, however, ties in ranking often do occur, and a special variation of tau can be used. It is

(6) $$\tau_b = \frac{P - Q}{\left[\left(\frac{n(n-1)}{2} - T_x\right) \cdot \left(\frac{n(n-1)}{2} - T_y\right)\right]^{1/2}}$$

where T_x is the number of pairs in which x scores are tied, T_y is the number of pairs in which y scores are tied, and T_{xy} is the number of pairs in which both the x and y scores are tied. Note that

(7) $$\binom{n}{2} = P + Q + T_x + T_y - T_{xy}$$

For computation of τ_b, consider the SAT scores for the 24 students shown in Table 4. Note that there are $24 \cdot 23/2 = 276$ different pairs of students, and there are four tied pairs in the x ranks, three from the 7.0's and one from the 15.5's. In general, if there are m cases with the same rank, these cases generate $\frac{m(m-1)}{2}$ tied pairs. Ties are not counted as correctly or incorrectly classified and hence are ignored when calculating P and Q. That is, when comparing students 13 and 12, the x and y rankings agree, but when comparing students 24 and 18, the order is neither correct nor incorrect. In this way, Q = 67 and P = 199, $T_x = 4$, $T_y = 7$, $T_{xy} = 1$.

$$\tau_b = \frac{199 - 67}{[(276 - 4)(276 - 7)]^{1/2}} = \frac{132}{(272 \cdot 269)^{1/2}} = \frac{132}{270.5} = .488$$

See Kendall (1970) and Goodman and Kruskal (1972) for significance tests of τ_b.

For data with tied observations, Kendall also introduces τ_a, the average of the values of τ obtained by all possible different rankings of tied observations. A variation of τ_a called τ_c can be used for contingency tables where the number of rows is large relative to the number of columns or vice versa.

Summary

Three measures of association for use when both variables are continuous have been discussed: Pearson's r, Spearman's rank correlation coefficient, and Kendall's tau. Pearson's r is a measure of linear association and requires the assumption of bivariate normality for valid tests of significance. Spearman's r_s and Kendall's τ measure the degree of monotone association (or linear association of ranks). For the data in Table 1, $r = -.92$, $r_s = -.95$, $\tau = -.86$. For the data in Table 2, $r = .69$, $r_s = .64$, $\tau_b = .49$.

The measures r_s and τ may be used when the original scores have resulted simply from two different rank orderings; measurements on an interval or ratio scale are unnecessary.

Categorical Variables

Association measures designed for the case where both variables are categorical include Goodman and Kruskal's γ, Somers' d, Lambda (λ), the uncertainty coefficient u, and a variety of measures based on χ^2 (chi square).

When both variables are categorical, the data is usually displayed in a classification table, such as Table 5. In this hypothetical example, two teachers were asked to estimate the achievement potential for 78 tenth-grade students who had signed up for a French language course. Their ratings were confined to the three categories of below average, average, and above average.

TABLE 5

Agreement on French Achievement Potential by Two Classroom Teachers

		Teacher 1 = y			
		Below average	Average	Above average	Totals
Teacher 2 = x	Below average	12	5	2	19
	Average	3	34	6	43
	Above average	1	7	8	16
	Totals	16	46	16	78

The values of the x variable, ratings by Teacher 2, are used to define the rows of the table, and the values of the y variable to define the columns. The number of individuals receiving score number i on the x variable and score number j on the y variable is entered in the ijth "cell" of the table and denoted by f_{ij}. Thus, the frequency of individuals rated as having average potential by Teacher 2 and as having above-average potential by Teacher 1 is $f_{23} = 6$. Row totals are denoted by $f_{i.}$ and column totals by $f_{.j}$; the grand total is $f_{..} = n$. For the data of Table 5, $f_{.2} = 46$ and $n = 78$.

Categorical variables may be ordered as in Table 5 or unordered as in variables such as field of study (major field). Some measures, such as γ, are useful when categories are ordered; others, like those based on χ^2 and λ, ignore any ordering of categories. For situations in which both categories of cross-classification are ordered, measures of association designed for continuous measures could be used by assigning scores to each category and using a procedure which handles ties well.

<u>Gamma</u>

The measure γ is basically a version of Kendall's τ developed for the case where the number of tied observations is large. This is the situation in a typical contingency table. For example, in Table 5 there are only three possible rankings for 78 students, and therefore it is impossible for each to have his own distinct rank. In fact, a great number are put in the same category by any one rater and are, hence, "tied." Gamma was developed by Goodman and Kruskal (1954) as a symmetric measure of monotone association for ordered categorical variables. Gamma is estimated by

(8) $$G = \frac{P - Q}{P + Q}$$

where P and Q have the same meaning as for Kendall's tau. When there are no ties, $P + Q = (n(n-1))/2$ and $G = \tau$. Note that tied observations are not considered in the calculation of G.

Figure 6 a, b, and c gives examples of tables that produce perfect monotone association (G = 1). Tables with patterns like these with relatively low frequencies in the "zero cells" will lead to high values of G. The value of G is zero in the table in Figure 6d in which the pattern of nonzero cells is not monotone.

f_{11}	0	0
0	f_{22}	0
0	0	f_{33}

(a) G = 1.0

f_{11}	f_{12}	0
0	f_{22}	f_{23}
0	0	f_{33}

(b) G = 1.0

f_{11}	0	0
f_{21}	0	0
f_{31}	f_{32}	f_{33}

(c) G = 1.0

f	0	f
0	f	0
f	0	f

(d) G = 0.0

Fig. 6. Tables illustrating various values of G. (Note that all cells marked f in table d have the same frequency.)

The continuous SAT data of Table 4, where the number of ties is small, yields

$$G = \frac{199 - 67}{199 + 67} = \frac{132}{266} = .496$$

which is close to τ_b = .521.

An algorithm for the calculation of G in contingency tables is illustrated for the data of Table 5. To get P, multiply every cell frequency f_{ij} by the sum of cell frequencies for all cells which lie to the right of and below that cell, and add these results. For Q, the same procedure is applied to cells to the left and below.

$$P = 12(34 + 6 + 7 + 8) + 3(7 + 8) + 5(6 + 8) + 34(8)$$

$$Q = 5(3 + 1) + 34(1) + 2(3 + 34 + 1 + 7) + 6(1 + 7)$$

$$P = 1047$$

$$Q = 192$$

$$G = \frac{1047 - 192}{1047 + 192} = \frac{855}{1239} = .690$$

Simple inspection of the data matrix in Table 5 indicates monotone agreement between the two teachers. In this situation, τ_b = .484, demonstrating how different G and τ_b may be when the number of ties is large.

Asymptotic tests and confidence intervals for the value of γ may be based on the asymptotic normal distribution of G. For details, see Goodman and Kruskal (1963, 1972); these results are not included here because of the tedious calculations involved. However, conservative asymptotic procedures for the case where the observations constitute a simple random sample from the population may be based on the assumption that G has approximately a normal distribution with mean γ and estimated variance

(9) $$s_G^2 = \frac{2n(1 - G^2)}{P + Q}$$

where G is given by (7).

Somers' d

Closely related to γ and τ are the asymmetric measures developed by Somers (1962). When predicting y from x is of interest,

(10) $$d_{yx} = \frac{P - Q}{P + Q + T_y - T_{xy}}$$

where P and Q are the same as for Kendall's tau, T_y is the number of pairs tied in y, and T_{xy} is the number of pairs in which both x and y are tied.

A d for the symmetric case given by Anderson and Zelditch (1968) is exactly the same as τ_b.

(11) $$d = \frac{P - Q}{[(P + Q + T_y - T_{xy})(P + Q + T_x - T_{xy})]^{1/2}}$$

The relation is $\tau_b^2 = d_{yx}d_{xy} = d^2$. Remember that $\binom{n}{2} = P + Q + T_x + T_y - T_{xy}$. Significance tests are based on $S = P - Q$ as for τ; see also Goodman and Kruskal (1972).

Lambda

The λ measures, both symmetric and asymmetric, were developed by Goodman and Kruskal (1954) as measures of predictive association, different in concept from the measures of monotone or linear association discussed so far. The basic purpose of λ_y is to measure the degree of success with which an individual's x value may be used to predict his y value. The prediction procedure when x is known is to pick the y value that has the highest frequency for that value of x; no use is made of any ordering of the actual score values. Using such a procedure one can define the probability of making an error in prediction when x is known, P(error|x known), and the probability of making an error in prediction when x is unknown and the value of y with the highest frequency overall is predicted. Then λ_y is defined as the proportionate reduction in the probability of making an error owing to the knowledge of x.

(12) $$\lambda_y = \frac{P(\text{error}|x \text{ unknown}) - P(\text{error}|x \text{ known})}{P(\text{error}|x \text{ unknown})}$$

To calculate L_y from a sample, the following argument is used. Suppose the relationship between x and y in the sample has been observed as in Table 5. We are then asked to guess the y score of an individual selected at random from the n individuals represented in the table. Without knowing that individual's x score, the best procedure is to predict the y value with the highest frequency. Define m as the index for which $f_{\cdot j}$ is a maximum,

$$\max_j f_{\cdot j} = f_{\cdot m}$$

and then predict the score value y corresponding to the index m. For the data of Table 5, suppose we consider using Teacher 2's assessment to predict Teacher 1's assessment. Without knowing Teacher 2's rating, we would predict that Teacher 1 would classify a student as average since $\max_j f_{\cdot j} = 46$, which occurs for $j = 2$. The number of prediction errors made by using this procedure is

(13) $$\text{Number of errors when x unknown} = n - \max_j f_{\cdot j} = n - f_{\cdot m}$$

Suppose, however, that we are allowed to utilize the individual's x score in making the prediction. If x has index i, we predict the score value of y corresponding to $\max_j f_{ij}$; that is, pick the value of y with the highest frequency for that value of x. Again define

$$f_{im} = \max_j f_{ij}$$

then, for the data in Table 5, $\max_j f_{1j} = 12$, $\max_j f_{2j} = 34$, $\max_j f_{3j} = 8$. The number of errors made using this procedure is $n - \sum_i f_{im}$, which is $78 - (12 + 34 + 8) = 24$ for the data in Table 5. Then L_y is the proportionate reduction in the number of errors when x is utilized.

$$(14)\quad L_y = \frac{\text{Number of errors when x is unknown} - \text{Number of errors when x is known}}{\text{Number of errors when x is unknown}}$$

$$= \frac{n - \max_j f_{\cdot j} - [n - \sum_i \max_j f_{\cdot j}]}{n - \max_j f_{\cdot j}}$$

$$= \frac{\sum_i f_{im} - f_{\cdot m}}{n - f_{\cdot m}}$$

where $f_{\cdot m}$ is the largest frequency from the column totals and $\sum_i f_{im}$ is the sum over all rows of the largest cell frequency in each row.

For the data in Table 5, $L_y = (12 + 34 + 8 - 46)/(78 - 46) = 8/32 = .25$. Note that L_y makes no use of any ordering of the actual score values of either x or y so it does not measure monotone association.

Tables such as those shown in Figure 7 illustrate the kinds of association yielding large or small L_y. Note that in general the values of L_y may be expected to be quite different from those of other measures of association. (For example, tables c and f in Figure 7 yield G = 1 and table d in Figure 7 yields G = 0).

f_{11}	0	0
0	0	f_{23}
0	f_{32}	0

(a) $L_y = 1$

f_{11}	0	0
0	f_{22}	0
0	0	f_{33}

(b) $L_y = 1$

f	f	0
0	f	f
0	0	f

(c) $L_y = 1/3$

f	0	f
0	f	0
f	0	f

(d) $L_y = 1/3$

f_{11}	0	0
f_{21}	0	0
f_{31}	0	0

(e) $L_y = 0$

f	0	0
f	0	0
f	f	f

(f) $L_y = 0$

Fig. 7. Tables illustrating various values of L_y. (All cells marked f have the same frequency.)

For predicting x from y, L_x is defined in an analogous manner. When a symmetric measure of association is desired (the direction of the prediction is not important), the composite measure L can be defined

$$\text{(15)} \qquad L = \frac{\sum_j \max_i f_{ij} + \sum_i \max_j f_{ij} - \max_i f_{i\cdot} - \max_j f_{\cdot j}}{2n - \max_i f_{i\cdot} - \max_j f_{\cdot j}}$$

Calculating L from the data of Table 5 yields these results:

$$\sum_j \max_i f_{ij} = 12 + 34 + 8 = 54$$

$$\sum_i \max_j f_{ij} = 12 + 34 + 8 = 54$$

$$\max_i f_{i\cdot} = 43$$

$$\max_j f_{\cdot j} = 46$$

$$L = \frac{54 + 54 - 43 - 46}{2(78) - 43 - 46} = \frac{19}{67} = .284$$

Note the large difference between G = .7 and L = .3. Because L is not a monotonicity measure like G but rather an index of predictive association, there is in general no reason to expect similar results.

To make tests or confidence intervals on the value of λ_y, when the observations constitute a simple random sample, we may use the fact that L_y is asymptotically normal with mean λ_y and variance estimated by

$$\text{(16)} \qquad s^2_{L_y} = \frac{(n - \sum_i f_{im})(\sum_i f_{im} + f_{\cdot m} - 2 \sum^r f_{im})}{(n - f_{\cdot m})^3}$$

where $\sum^r f_{im}$ denotes summation of the maximum frequency in a row only over those rows in which f_{im} falls in the same column as $f_{\cdot m}$. So, for the data in Table 5, $f_{1m} = 12$, $f_{2m} = 34$, $f_{3m} = 8$, and the column in which $f_{\cdot m}$ occurs is column 2 ($f_{\cdot m} = 46$); thus, $\sum^r f_{im} = 34$.

These properties of asymptotic normality hold under the assumptions that a random sample of size n has been drawn from the population, that λ_y is not equal to zero or one, that in the population the maximum proportions p_{im} and $p_{\cdot m}$ are unique and $p_{\cdot m} \neq 1.0$. Significance tests and confidence intervals for λ_x and λ are derived in a similar fashion; see Goodman and Kruskal (1963, 1972) for complete details.

Uncertainty Coefficient

The uncertainty coefficient u, sometimes called the coefficient of constraint, was developed by information theorists as an asymmetric or symmetric measure of association based on the reduction of uncertainty about one variable when the other variable is known. Thus it is conceptually similar to λ.

An explanation for the asymmetric case is developed below. Suppose we want to predict the value of y. The "uncertainty" about an individual's y value when x is unknown depends on the marginal distribution of y and is defined as

$$(17) \qquad U(y) = -\sum_j \frac{f_{\cdot j}}{n} \log\left(\frac{f_{\cdot j}}{n}\right)$$

The base of the logarithm is arbitrary, but base two is frequently used following the lead of Claude Shannon, the information-theory pioneer, who originally defined (17) as a measure of entropy.

When x is known

$$(18) \qquad U(y|x) = -\sum_i \sum_j \frac{f_{ij}}{n} \log\left(\frac{f_{ij}}{f_{i\cdot}}\right)$$

The uncertainty coefficient denoted here by u_y when predicting y from x is the reduction in uncertainty due to knowledge of x:

$$(19) \qquad u_y = \frac{U(y) - U(y|x)}{U(y)}$$

Note the similarity to the definition of λ_y.

For the data in Table 5, using logarithms to the base 2 ,

$$U(y) = -\left[\frac{16}{78}\log\frac{16}{78} + \frac{46}{78}\log\frac{46}{78} + \frac{16}{78}\log\frac{16}{78}\right] = 1.3865$$

$$U(y|x) = -\left[\frac{12}{78}\log\frac{12}{19} + \frac{3}{78}\log\frac{3}{43} + \frac{1}{78}\log\frac{1}{16}\right.$$

$$+ \frac{5}{78}\log\frac{5}{19} + \frac{34}{78}\log\frac{34}{43} + \frac{7}{78}\log\frac{7}{16}$$

$$\left. + \frac{2}{78}\log\frac{2}{19} + \frac{6}{78}\log\frac{6}{43} + \frac{8}{78}\log\frac{8}{16}\right] = 1.0822$$

and $\qquad u_y = .219$

This value is reasonably close to $L_y = .25$ for the same data.

A symmetric version of the uncertainty coefficient is

$$(20) \qquad u = \frac{U(y) + U(x) - U(y,x)}{U(y) + U(x)}$$

where

$$(21) \qquad U(y,x) = -\sum_i \sum_j \frac{f_{ij}}{n} \log \left(\frac{f_{ij}}{n}\right)$$

For more information and significance tests, see Attneave (1959).

Chi Square

The chi-square test is commonly used to test for association between categorical variables. The chi-square test ignores ordering of the variables; that is, it is insensitive to the score values of x and y. The χ^2 statistic itself cannot be used as a measure of association since it ranges from 0 to ∞; a number of transformations of χ^2 have been proposed to serve as association measures.

For a test of independence between two categorical variables

$$(22) \qquad \chi^2 = \sum_{ij}\sum \frac{n(f_{ij} - f_{i.}f_{.j}/n)^2}{f_{i.}f_{.j}}$$

For the data of Table 5, $\chi^2 = 37.87$ with 4 degrees of freedom; see the computations in Table 6.

TABLE 6

Calculation of Chi Square for Data from Table 5

Cell	Observed frequencies (f_{ij})	Expected frequencies $\left(\frac{f_{i.}f_{.j}}{n}\right)$	$\frac{n(f_{ij} - f_{i.}f_{.j}/n)^2}{f_{i.}f_{.j}}$
1-1	12	3.90	16.82
1-2	5	11.21	3.44
1-3	2	3.90	.93
2-1	3	8.82	3.84
2-2	34	25.36	2.94
2-3	6	8.82	.90
3-1	1	3.28	1.58
3-2	7	9.44	.63
3-3	8	3.28	6.79

Any table in which the cell frequencies differ markedly from the frequencies to be expected if the variables were independent can lead to a large value of χ^2. That is, a large χ^2 does not imply the existence of any particular type of association. Only a table such as that shown in Table 7 can lead to a zero value of χ^2.

TABLE 7

A Contingency Table in Which $\chi^2 = 0$

12	12	6	30
20	20	10	50
8	8	4	20
40	40	20	100

When n is large and none of the expected cell frequencies is too small (see Cochran, 1954), the calculated χ^2 can be referred to a table of the χ^2 distribution using degrees of freedom = $(r - 1)(c - 1)$ to test the null hypothesis of independence or no association between the variables of classification. (Here r is the number of rows and c the number of columns.)

A substitute for the standard χ^2 test that is less sensitive to small frequencies is the method based on the likelihood ratio, which is given in Mood and Graybill (1963). For 2x2 tables with small n, Fisher's exact test can be used for significance testing (see Siegel, 1956).

There is one basic transformation for turning a computed χ^2 statistic into an index of association ranging from 0 to 1. For a 2x2 contingency table, this index is called phi, ϕ, and for a larger table it is called Cramer's V. It is

$$V = \left[\frac{\chi^2}{n(m - 1)}\right]^{1/2} \tag{23}$$

where m is the number of rows or the number of columns, whichever is smaller. For the data of Table 5, V = .49. Another transformation of χ^2 called the contingency coefficient has been proposed

$$(24) \qquad C = \left[\frac{\chi^2}{\chi^2 + n}\right]^{1/2}$$

However, since C has a maximum value less than 1.0 dependent on r and c, it is not widely used.

When there are two rows and two columns, $m - 1 = 1$ and $V = (\chi^2/n)^{1/2}$. In this special case the measure of association is denoted by ϕ.

$$(25) \qquad \phi = \sqrt{\frac{\chi^2}{n}}$$

The data shown in Table 8 may be used to illustrate the use of ϕ. A junior high school principal was interested in a quick index of association between academic achievement and musical inclination for the pupils in his school. Although both are continuous variables, he forced each into two easily observed categories, namely on "honor roll" or "not on honor roll" and "music interest" or "no music interest" (participation in band or glee club or two music courses constituted "interest"). For this data, $\chi^2 = 98.8$, $\phi = .393$.

TABLE 8

Six Hundred Forty Junior High School Students
Classified on Academic Achievement and "Music Interest"

	Honor Roll	Not Honor Roll	
Music	76	42	118
No Music	100	422	522
	176	464	640

An alternate formula for ϕ can be derived from Pearson's r by assigning values of 0 and 1.0 to the row and column categories. The results of the two approaches will be identical. Unfortunately ϕ can be as great as 1.0 only when the marginal frequencies are the same for both variables. See Guilford (1965) or Walker and Lev (1953) for further discussions of ϕ.

Tetrachoric r

This measure is an approximation to Pearson's r for continuous variables x and y, which have both been forced into dichotomies to yield a 2x2 contingency table. The computation formula for r is long and complicated. One approximation formula is

$$r_t = \cos\left(\frac{180° \sqrt{f_{12}f_{21}}}{\sqrt{f_{11}f_{22}} + \sqrt{f_{12}f_{21}}}\right) \tag{26}$$

Tables in Guilford (1965) can be entered with $f_{11}f_{22}/f_{12}f_{21}$ to give the value of r_t computed from formula (26). This formula only provides a good approximation when both variables are split at their medians. For the data in Table 8, $r_t = .671$. Note the wide divergence between the values obtained for r_t and ϕ even though both measures are based on r. For a complete explanation of tetrachoric r and information on significance testing, see Guilford (1965).

Summary

When both variables are categorical and both are ordered, Goodman and Kruskal's γ or Somers' d provide measures of association based on counting agreements and disagreements between x and y rankings for all pairs of observations. When either the x or the y variable is not ordered or nonmonotone relationships between x and y are of interest, the asymmetric or symmetric measures of predictive association λ, or of uncertainty u, or a measure based on Pearson's r or χ^2 may be used. For any of these measures, significance tests of the null hypothesis of no association of the specified type rely on large-sample normal or χ^2 approximations.

For the data in Table 5, $\gamma = .59$, $d = .48$, $\lambda = .28$, $u = .11$, $v = .49$ and $d_{yx} = .47$, $\lambda_y = .25$, $u_y = .22$.

One Continuous and One Categorical Variable

For the situation where one variable is continuous and the other categorical, we denote the continuous variable as y and the categorical variable as x. No measures of association have been developed specifically for the case where x is ordered and multicategorical. Thus, two options are available if the categorical variable is ordered: ordered scores could be assigned to x and a measure for continuous variable such as r, r_s, or τ computed; or, y could be categorized and γ computed.

In this section we review measures based on the correlation ratio, η^2, for the case where x has several categories, point-biserial and biserial r, and modifications of γ and λ for use when x is dichotomous. Table 9 constitutes a hypothetical example; it represents Reading Readiness raw scores (instead of the customary percentiles) from a 0 to 140 scale for 30 children with kindergarten training.

TABLE 9

Reading Readiness Raw Scores for Thirty Children with Kindergarten Training

Male	Female
62	66
63	68
68	68
70	70
72	75
72	77
73	77
74	78
76	79
80	80
82	81
85	83
86	85
89	86
	87
	89

$n_M = 14$	$n_F = 16$	$n = 30$
$\overline{M} = 75.14$	$\overline{F} = 78.06$	$\overline{T} = 76.70$
$\sigma_M = 8.33$	$\sigma_F = 7.19$	$\sigma_T = 7.74$

Correlation Ratio

The correlation ratio based on a continuous y variable and a categorical x variable is

$$(27) \qquad \eta^2_{yx} = \frac{\sum_i n_i (\overline{y}_{i.} - \overline{y}_{..})^2}{\sum_i \sum_j (\overline{y}_{ij} - \overline{y}_{..})^2}$$

where j is the index for the grouped variable, $\overline{y}_{i.}$ is the mean of the y scores for individuals with x = i, and $\overline{y}_{..}$ is the mean of all the y scores. In analysis of variance terms

$$(28) \qquad \eta^2_{yx} = \frac{\text{Sum of squares between groups}}{\text{Sum of squares total}}$$

Thus η^2_{yx} can be interpreted as the proportion of variance in the y variable which can be accounted for by knowing the classification of the x variable. Note that the association between the means of y and the classification x may be of any type. For the data shown in Table 9, $\eta^2_{yx} = .02$. For details see Hays (1963) or Guilford (1965).

Other measures of association closely related to η^2 are ω^2, ε^2, and ρ_i, the intraclass correlation. See Glass and Hakstian (1969) for a comparative discussion of these measures and their interpretation. The significance test for η^2 and the related measures is simply the standard F test.

Biserial and Point-Biserial Correlation

When one variable is continuous and the other dichotomous, the traditional measures of association have been point-biserial, r_{pb}, and biserial, r_b. They are based on Pearson's r and assume normality of the continuous variable. The measures r_{pb} and r_b will approach 1.0 as the means of the y variable in the two x categories get farther and farther apart relative to the within-group standard deviation. The measures will be zero when the mean of the y variable is the same for both values of x. For a fixed difference in the y means in the two groups relative to the within group standard deviation, the value of r_{pb} varies with the quantity of n_1n_2/N^2; that is, r_{pb} will be larger for $n_1 = n_2$ than for $n_1 \neq n_2$ for fixed N.

Suppose a measure of association between sex and reading readiness was desired. For convenience, the data of Table 9 have been arranged in ascending order. The point-biserial r is simply Pearson's r evaluated with labels 1 and 2 assigned to the two x categories; a simple computing formula is the following:

$$(29) \qquad r_{pb} = \frac{(\bar{y}_2 - \bar{y}_1)}{s_t} \sqrt{\frac{n_1 n_2}{N^2}}$$

where

$\bar{y}_1$ = mean of continuous variable in category x = 1

$\bar{y}_2$ = mean of continuous variable in category x = 2

n_1 = number of cases in category x = 1

n_2 = number of cases in category x = 2

$N = n_1 + n_2$

s_t = total standard deviation of the continuous variable.

For the data of Table 9, let girls constitute category 2 and boys category 1, to obtain

$$r_{pb} = \frac{\bar{y}_2 - \bar{y}_1}{s_t} \sqrt{\frac{n_1 n_2}{N^2}}$$

$$= \frac{78.06 - 75.14}{7.74} \sqrt{(.467)(.533)}$$

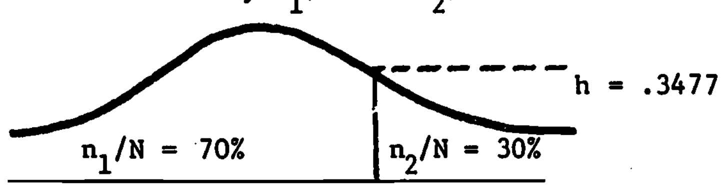
$$= .188$$

Conditional on the observed values of n_1 and n_2, a test of the null hypothesis of no association may be performed just as for Pearson's r under the assumption that the variable y has a normal distribution with the same variance in each category of x.

If the x variable was formed by collapsing a continuous variable with a normal distribution into a dichotomy, the biserial correlation coefficient may be used:

$$r_b = \left[\frac{\bar{y}_2 - \bar{y}_1}{s_t}\right] \cdot \frac{n_1 n_2}{hN^2} \tag{30}$$

where h is the ordinate from the unit normal distribution, read at the point where the areas defined by n_1/N and n_2/N meet.



Significance tests on the biserial correlation coefficient are usually done in the same manner as for Pearson's r; see Guilford (1965) for a complete discussion.

Gamma$_D$

Elashoff (1971) developed a version of Goodman and Kruskal's γ for the case of a dichotomous and a continuous variable.

(31) $$G_D = \frac{2U}{n_1 n_2} - 1$$

where U is the Mann-Whitney U statistic. Consider all pairs of observations in which one member of the pair belongs in category 1 and the other in category 2: there are $n_1 n_2$ pairs; then, U is the number of pairs in which the y observation from category 2 exceeds the y observation from category 1. The U statistic can also be calculated from the relationship

(32) $$U = n_1 n_2 + n_1(n_1 + 1)/2 - R_1$$

where all the y observations are ranked from 1 to $n_1 n_2$ and R_1 is the sum of the ranks for observations in category 1.

The measure G_D will get larger as the medians of the y observations in the two categories get farther apart relative to the within-group ranges.

For the data in Table 9, the ranks of the observations are

$$\text{Males} = \{1, 2, 5.0, 7.5, 9.5, 9.5, 11, 12, 14, 19.5, 22, 24.5, 26.5, 29.5\}$$

$$\text{Females} = \{3, 5.0, 5.0, 7.5, 13, 15.5, 15.5, 17, 18, 19.5, 21, 23, 24.5, 26.5, 28, 29.5\}$$

$$R_M = 193.5$$

$$U = 135.5$$

$$G_D = \frac{2(145.5)}{224} - 1 = \frac{271}{224} - 1 = 1.21 - 1.0 = .21$$

Conditional on n_1 and n_2, and assuming that the y variable is continuous, a test of the null hypothesis that the distributions of y in the two categories of x are identical may be made using tables and tests for the U statistic or the Wilcoxon rank-sum statistics R_1 or R_2; see, e.g., Dixon and Massey (1969). If the number of ties is small, the midrank procedure shown above may be used. If the number of ties is large, the y variable should probably be categorized and a measure for categorical variables used.

Lambda Measures

Several measures of predictive association related to Goodman and Kruskal's λ_x were developed by Elashoff (1971) for use when prediction of the dichotomous variable x is of interest.

A measure sensitive to differences in mean of the continuous variable corresponding to the two values of x is

$$(33) \qquad \lambda_1 = 1 - \frac{P(\text{ misclassification} \mid y \text{ known })}{P(\text{ misclassification} \mid y \text{ unknown })}$$

The calculation of λ_1 is based on the cumulative frequency distributions of y for each x category. If we label the dichotomous categories 1 and 2, as usual, then define

$$d^+ = \max_a \, (N_1(a) - N_2(a))$$

$$d^- = \max_a \, (N_2(a) - N_1(a))$$

$$d = \max \, (d^+, d^-)$$

where $N_i(a)$ represents the cumulative frequency of y scores less than or equal to a in category i. If $d^+ > d^-$, then

$$(34) \qquad \lambda_1 = \frac{\dfrac{n_2 + d}{N} - m}{1 - m}$$

where $m = \max\{n_1/N, n_2/N\}$.

TABLE 10

Cumulative Frequency Distributions for Reading Readiness Scores

Scale point	N_1 (male)	N_2 (female)	Difference ($N_2 - N_1$)
62	1	0	-1
63	2	0	-2
66	2	1	-1
68	3	3	0
70	4	4	0
72	6	4	-2
73	7	4	-3
74	8	4	-4
75	8	5	-3
76	9	5	-4
77	9	7	-2
78	9	8	-1
79	9	9	0
80	10	10	0
81	10	11	1
82	11	11	0
83	11	12	1
85	12	13	1
86	13	14	1
87	13	15	2
89	14	16	2

If $d^+ < d^-$, then

$$(35) \qquad \lambda_1 = \frac{\frac{n_1 + d}{N} - m}{1 - m}$$

This measure is based on the procedure of classifying an individual into category 1 if $y \leq a^*$ and into category 2 otherwise, where a^* is the scale point at which the value d occurs.

Table 10 shows the cumulative frequency distributions for the data of Table 9, along with the d^+ and d^- calculations.

Since $d^- = 4$ is greater than $d^+ = 2$, we have

$$\lambda_1 = \frac{\frac{14 + 4}{30} - \frac{16}{30}}{1 - \frac{16}{30}} = \frac{2}{14} = .143$$

See Elashoff (1971) for further discussion and significance tests. In the situation in which the variances may differ, a λ_2 statistic has been defined in Elashoff (1971).

Summary

No measures of association have been developed for the case where one variable is continuous and the other ordered categorical. When the categorical variable is not ordered, a measure related to the correlation ratio which gives the proportion of variance in y accounted for by x may be applied. When x is dichotomous, two variants of Pearson's r, r_{pb} and r_b, are available, and variants of γ and λ_x have been developed.

References

Anderson, T. R., & Zelditch, M. A basic course in statistics with sociological applications. New York: Holt, Rinehart & Winston, 1968.

Attneave, F. Applications of information theory to psychology. New York: Holt, Rinehart & Winston, 1959.

Blalock, H. M., Jr. Social statistics. New York: McGraw-Hill, 1960.

Bradley, J. V. Distribution-free statistical tests. New York: Prentice-Hall, 1968.

Carroll, J. B., The nature of the data, or How to choose a correlation coefficient. Psychometrika, 1961, 26, 347-72.

Cochran, W. G. Some methods for strengthening the common χ^2 tests. Biometrics, 1954, 10, 417-51.

Dixon, W. J., & Massey, F. J., Jr. Introduction to statistical analysis. (3rd ed.) New York: McGraw-Hill, 1969.

Das Gupta, S. Non-parametric classification rules. Sankhya, Series A, 26, 25-30.

Elashoff, J. D. Measures of association between a dichotomous and a continuous variable. Proceedings of the American Statistical Association: Social Statistics Section, 1971, 298-300.

Fisher, R. A. Statistical methods for research workers. (14th ed.) Edinburgh: Oliver & Boyd, 1970.

Glass, G. V., & Hakstian, A. R. Measures of association in comparative experiments: Their development and interpretation. American Educational Research Journal, 1969, 6, 403-14.

Gnedenko, B. V. Kriterien für die Unveränderlichkeit der Wahrscheinlichkeitsverteilung von zwei unabhängigen Stichprobenreihen. Mathematisches Nachrichten, 1954, 12, 29-66.

Goodman, L. A., & Kruskal, W. H. Measures of association for cross classifications. Journal of the American Statistical Association, 1954, 49, 732-64.

Goodman, L. A., & Kruskal, W. H. Measures of association for cross classifications. II: Further discussion and references. Journal of the American Statistical Association, 1959, 54, 123-63.

Goodman, L. A., & Kruskal, W. H. Measures of association for cross classifications. III: Approximate sampling theory. Journal of the American Statistical Association, 1963, 58, 310-64.

Goodman, L. A., & Kruskal, W. H. Measures of association for cross classifications. IV: Simplification of Asymptotic Variances. Journal of the American Statistical Association, 1972, 67, 415-21.

Guilford, J. P. Fundamental statistics in psychology and education. New York: McGraw-Hill, 1965.

Hays, W. L. Statistics for psychologists. New York: Holt, Rinehart & Winston, 1963.

Kendall, M. G. Rank correlation methods. (4th ed.) New York: Hafner, 1970.

Kruskal, W. H. Ordinal measures of association. Journal of the American Statistical Association, 1958, 53, 814-61.

Lancaster, H. O., & Hamdan, M. A. Estimation of the correlation coefficient in contingency tables with possibly nonmetrical characters. Psychometrika, 1964, 29, 383-91.

Meredith, W. M. Basic mathematical and statistical tables for psychology and education. New York: McGraw-Hill, 1967.

Mood, A. M., & Graybill, F. A. Introduction to the theory of statistics. (2nd ed.) New York: McGraw-Hill, 1963.

Mosteller, F. Association and estimation in contingency tables. Journal of the American Statistical Association, 1968, 63, 1-28.

Rosenthal, I. Distribution of the sample version of the measure of association, gamma. Journal of the American Statistical Association, 1966, 61, 440-53.

Siegel, S. Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill, 1956.

Somers, R. H. A new asymmetrical measure of association for ordinal variables. American Sociological Review, 1962, 27, 799-811.

Stoller, D. S. Univariate two-population distribution-free discrimination. Journal of the American Statistical Association, 1954, 48, 770-79.

Tate, R. F. Applications of correlational model for biserial data. Journal of the American Statistical Association, 1955, 50, 1078-95.

Treloar, A. E. Correlation analysis. Minneapolis, Minn.: Burgess, 1942.

Walker, H. M., & Lev, J. Statistical inference. New York: Holt, Rinehart & Winston, 1953.

Walker, H. M., & Lev, J. Elementary statistical methods. (3rd ed.) New York: Holt, Rinehart & Winston, 1969.

Walsh, J. E. Handbook of nonparametric statistics. Princeton, N. J.: Van Nostrand, 1962. 3 vols.